# STABILITY CONTROL IN A TWO-CHANNEL SPEECH REINFORCEMENT SYSTEM FOR VEHICLES

*Alfonso Ortega, Eduardo Lleida, Enrique Masgrau, Luis Buera and Antonio Miguel**

Communication Technologies Group (GTC)
University of Zaragoza, Spain
{ortega,lleida,masgrau,lbuera,amiguel}@unizar.es

## ABSTRACT

This paper presents a two-channel speech reinforcement system for cars able to improve the communication between the front and the rear passengers. One of the problems of this kind of systems is that they must operate in closed-loop, as acoustic feedback paths appear due to the short distance between loudspeakers and microphones. This feedback paths can make the system become unstable and acoustic echo control is needed in order to ensure stability. The system must perform two plant identifications for each channel. One of them is an open-loop identification and the other one is closed-loop. We propose here the use of echo suppression filters specially designed for closed-loop subsystems along with echo suppression filters for open-loop subsystems based on the optimal filtering theory. Results about the performance of the proposed system are provided.

## 1. INTRODUCTION

Inside a car, speech intelligibility can be degraded due to the distance between speakers and the level of noise among other factors. Using a set of microphones placed on the ceiling of the cabin, a speech reinforcement system picks up the speech of each passenger, amplifies and plays those signals back into the cabin using the loudspeakers of the car audio system [1].

Acoustic echo appears because the signal radiated by the loudspeakers is picked up again by the microphones. Due to the amplification stage between the microphones and the loudspeakers, the system can become unstable.

Along with the speech signal, the noise is also picked up by the microphones and amplified by the system increasing the overall noise level present inside the car. To prevent this, Noise Reduction Filters (NRF) must be used.

According to Fig. 1, in a two channel speech reinforcement system, we can identify two closed-loop subsystems, one for each channel, and two open-loop subsystems.

Acoustic Echo Cancellers (AEC) are widely used to overcome electro-acoustic coupling between loudspeakers and microphones. Nevertheless, the use of Echo Suppression

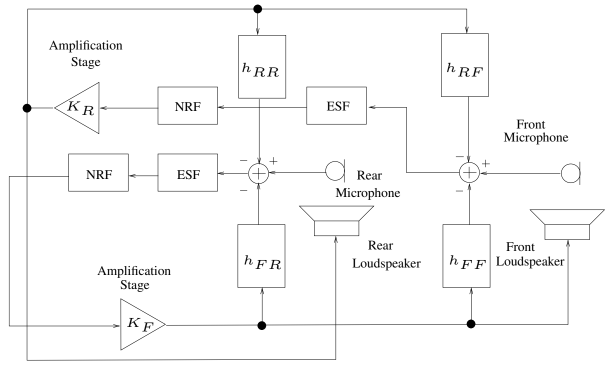


**Fig. 1**. *Schematic diagram of a two-channel speech reinforcement system for cars.*

Filters (ESF), along with acoustic echo cancellers, to achieve enough echo attenuation is presented here. Several techniques have been proposed for further echo attenuation using residual echo reduction filters [2, 3]. These techniques can be used for open-loop systems but in a speech reinforcement system for vehicles, due to its closed-loop operation, the ESF must also ensure stability. The study for a one-channel system can be found in [4]. In this paper, the optimal ESF transfer function for the closed-loop subsystems in a two-channel speech reinforcement system is derived.

Another important aspect of this system is that the overall delay must be short enough to achieve full integration of the sound coming from the direct path and the reinforced speech coming from the loudspeakers.

This paper is organized as follows. A brief description and a stability study of the system will be presented in Section 2, along with the optimal expressions for the Echo Suppression Filters in the two-channel system. In Section 3, the proposed Echo Suppression Filters will be presented. In Section 4, performance measures and results will be shown and in Section 5, we present the conclusion along with a summary of the paper.

## 2. DESCRIPTION AND STABILITY OF THE TWO-CHANNEL SYSTEM

We can describe the two channel system by using the relationships between each input-output pair, according to

* This work was funded by the grant TIC2002-04103-C03-01 from the Spanish MEC.

$$\begin{bmatrix} T_{RF}(e^{j\omega}) & T_{RR}(e^{j\omega}) \\ T_{FF}(e^{j\omega}) & T_{FR}(e^{j\omega}) \end{bmatrix} \begin{bmatrix} S_F(e^{j\omega}) \\ S_R(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} O_R(e^{j\omega}) \\ O_F(e^{j\omega}) \end{bmatrix}, \quad (1)$$

where $T_{XY}(e^{j\omega})$ is the transfer function that relates the output from loudspeaker X to the input of the microphone Y, $O_X(e^{j\omega})$ is the output from loudspeaker X and $S_Y(e^{j\omega})$ is the input corresponding to microphone Y, being X and Y, R for rear or F for front.

According to Fig. 1, we can express each output-input transfer functions as

$$T_{XY}(e^{j\omega}) = \frac{K_X W_X(e^{j\omega})\left(1 - K_Y W_Y(e^{j\omega})\tilde{H}_{YX}(e^{j\omega})\right)}{D(e^{j\omega})} \quad (2)$$

$$T_{XX}(e^{j\omega}) = \frac{K_X W_X(e^{j\omega}) K_Y W_Y(e^{j\omega})\tilde{H}_{YY}(e^{j\omega})}{D(e^{j\omega})}, \quad (3)$$

with

$$\begin{aligned} D(e^{j\omega}) &= \left[1 - K_F\tilde{H}_{FR}(e^{j\omega})\right]\left[1 - K_R\tilde{H}_{RF}(e^{j\omega})\right] \\ &\quad - K_R K_F \tilde{H}_{RR}(e^{j\omega})\tilde{H}_{FF}(e^{j\omega}), \end{aligned} \quad (4)$$

where $\tilde{H}_{XY}(e^{j\omega})$ is the difference between the Loudspeaker-Enclosure-Microphone (LEM) path transfer function, $H_{XY}(e^{j\omega})$, from loudspeaker $X$ to microphone $Y$, and its corresponding adaptive filter transfer function $\hat{H}_{XY}(e^{j\omega})$. $W_R(e^{j\omega})$ is the transfer function of the system composed of the ESF and NRF for the front-rear channel and $W_F(e^{j\omega})$ for the rear-front channel. $K_F$ and $K_R$ are the gain factors for the rear-front channel and the front-rear channel respectively.

The optimal transfer functions, that ensures unconditional stability and complete isolation between channels must satisfy

$$T_{XY}(e^{j\omega}) = K_X W_{X_n}(e^{j\omega}) \quad T_{XX}(e^{j\omega}) = 0, \quad (5)$$

where $W_{Fn}(e^{j\omega})$ and $W_{Rn}(e^{j\omega})$ are the transfer functions of the noise reduction filter of the rear-front channel and the front-rear channel respectively.

Substituting the conditions in (5) into (2), and considering (4), the optimal echo suppression filter expression for each channel is

$$W_{Xe}(e^{j\omega}) = \frac{W_{Ye}(e^{j\omega})}{D_{Xe}(e^{j\omega})}, \quad (6)$$

with

$$\begin{aligned} D_{Xe}(e^{j\omega}) &= 1 - K_Y W_{Yn}(e^{j\omega}) W_{Ye}(e^{j\omega})\tilde{H}_{YX}(e^{j\omega}) \\ &\quad + K_X W_{Ye}(e^{j\omega}) W_{Xn}(e^{j\omega})\tilde{H}_{XY}(e^{j\omega}), \end{aligned} \quad (7)$$

Expression in (6) shows that both ESF are not independent and their existence is only possible if

$$K_R\tilde{H}_{RF}(e^{j\omega}) = K_F\tilde{H}_{FR}(e^{j\omega}), \quad (8)$$

for each frequency, which implies that both filters must be equal to each other. Condition (8), is not under the control of the designer, so it will not be always met.

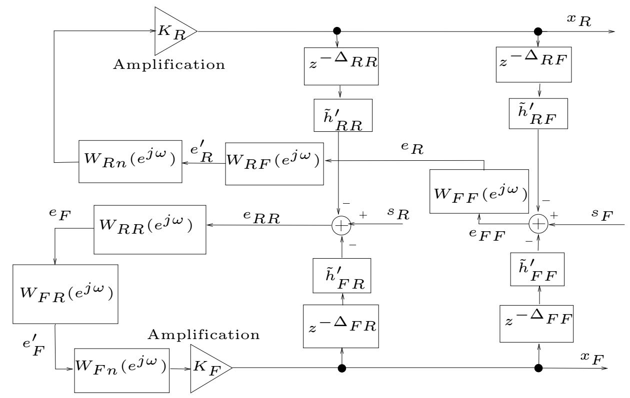


**Fig. 2**. *Two-channel speech reinforcement system with diferentiated treatment techniques for closed-loop subsystems and for open-loop subsystems.*

## 3. ECHO SUPPRESSION FILTERS FOR THE CLOSED-LOOP SUBSYSTEMS AND THE OPEN LOOP SUBSYSTEMS

One possible solution to increase the stability of the two-channel speech reinforcement system is to distinguish between open-loop subsystems and closed-loop subsystems applying specific treatment approaches to each one of them.

To cope with the residual echo remaining after the echo canceller for the open-loop subsystems, several approaches have been proposed in the literature [3]. The use of the filters $W_{FF}(e^{j\omega})$ and $W_{RR}(e^{j\omega})$, that follow a Wiener based approach, is considered.

In order to increase the stability margin of the speech reinforcement system, we propose here the use of the echo suppression filters $W_{RF}(e^{j\omega})$ and $W_{FR}(e^{j\omega})$, specially designed for the closed-loop subsystems.

The proposed system is presented in Fig. 2 where $s_R$ and $s_F$ are the input signals for the rear-front channel and the front-rear channel respectively, $x_R$ is the output signal of the front-rear channel and $x_F$ is the output signal of the rear-front channel. Due to the propagation delay, the LEM path of each loudspeaker-microphone pair is modeled as a delay block of $\Delta_{XY}$ samples ($X$ refers to the loudspeakers, front or rear, and $Y$ refers to the microphones, front or rear) followed by a linear system with the same impulse response of the LEM path except for the first $\Delta_{XY}$ values. The first $\Delta_{XY}$ coefficients of its corresponding adaptive filter are also set to zero to compensate for the propagation delay.

According to Fig. 2, and using

$$W_{RF}(e^{j\omega}) = \frac{1}{1 + K_R W_{FF}(e^{j\omega}) W_{Rn}(e^{j\omega})\tilde{H}_{RF}(e^{j\omega})} \quad (9)$$

$$W_{FR}(e^{j\omega}) = \frac{1}{1 + K_F W_{RR}(e^{j\omega}) W_{Fn}(e^{j\omega})\tilde{H}_{FR}(e^{j\omega})}, \quad (10)$$

as the proposed ESF for the closed-loop subsystems, the transfer functions for each input-output pair follow expressions (11) and (12).

$$T_{XY}(e^{j\omega}) = \frac{K_X W_{YY}(e^{j\omega}) W_{Xn}(e^{j\omega})}{1 - K_R K_F W_{FF}(e^{j\omega}) W_{RR}(e^{j\omega}) W_{Rn}(e^{j\omega}) W_{Fn}(e^{j\omega}) \tilde{H}_{FF}(e^{j\omega}) \tilde{H}_{RR}(e^{j\omega})} \quad (11)$$

$$T_{XX}(e^{j\omega}) = \frac{K_X W_{YY}(e^{j\omega}) W_{Xn}(e^{j\omega}) K_Y W_{XX}(e^{j\omega})) W_{Yn}(e^{j\omega}) \tilde{H}_{YY}(e^{j\omega})}{1 - K_R K_F W_{FF}(e^{j\omega}) W_{RR}(e^{j\omega}) W_{Rn}(e^{j\omega}) W_{Fn}(e^{j\omega}) \tilde{H}_{FF}(e^{j\omega}) \tilde{H}_{RR}(e^{j\omega})} \quad (12)$$

Thus, the stability of the reinforcement system, assuming that the echo suppression filters are working properly, depends only on the open-loop subsystems. That is, the stability depends on the misadjustment functions $\tilde{H}_{RR}(e^{j\omega})$ and $\tilde{H}_{FF}(e^{j\omega})$ that is intended to be minimized by the filters $W_{RR}(e^{j\omega})$ and $W_{FF}(e^{j\omega})$ respectively.

Each ESF, depends on the misadjustment functions of its corresponding closed-loop subsystems that are a priori unknown. Assuming that the ESF for the open-loop subsystems are real valued functions, as well as the NRF for each channel, it can be shown [1], that using the magnitude of the misadjustment function is the best option to increase the stability of the system. The estimates of the misadjustment function magnitude for each closed-loop subsystem are obtained using estimates of the residual echo $r_{FF}(n)$ for the rear-front channel and estimates of the residual echo $r_{RR}(n)$ for the front-rear channel, according to Fig.2.

For the front-rear channel, the residual echo remaining after the closed-loop subsystem acoustic echo canceller, can be expressed as

$$r_{RF}(n) = x_R(n) * w_{FF}(n) * \tilde{h}_{RF}(n), \quad (13)$$

where $w_{FF}(n)$ is the impulse response of the ESF for the open-loop subsystem of the front-rear channel and $\tilde{h}_{RF}(n)$ is the inverse Fourier transform of the misadjustment function. Thus, the Power Spectral Density (PSD) of the residual echo can be expressed as

$$S_{r_{RF}}(e^{j\omega}) = S_{x_R}(e^{j\omega}) \cdot \left| W_{FF}(e^{j\omega}) \tilde{H}_{RF}(e^{j\omega}) \right|^2, \quad (14)$$

which depends on the PSD of the output signal that will be played back through the rear loudspeakers of the reinforcement system, $S_{x_R}(e^{j\omega})$, and on the squared magnitude of the misadjustment function, $|\tilde{H}_{RF}(e^{j\omega})|^2$, along with the squared magnitude of the ESF of the open-loop subsystem of the front-rear channel, $\left|W_{FF}(e^{j\omega})\right|^2$.

The PSD of the rear output signal, according to Fig. 2, can be expressed as

$$S_{x_R}(e^{j\omega}) = S_{e_R}(e^{j\omega}) \cdot K_R^2 \cdot \left| W_{RFe}(e^{j\omega}) W_{Rn}(e^{j\omega}) \right|^2, \quad (15)$$

and thus, combining (14) and (15) and substituting into (9), we can obtain the expression for the closed-loop ESF of the front-rear channel that responds to

$$W_{RF}(e^{j\omega}) = 1 - \sqrt{\frac{S_{r_{RF}}(e^{j\omega})}{S_{e_R}(e^{j\omega})}}. \quad (16)$$

In the same way, we can obtain the expression for the ESF for the closed-loop subsystem of the rear-front channel that must follow

$$W_{FR}(e^{j\omega}) = 1 - \sqrt{\frac{S_{r_{FR}}(e^{j\omega})}{S_{e_F}(e^{j\omega})}}. \quad (17)$$

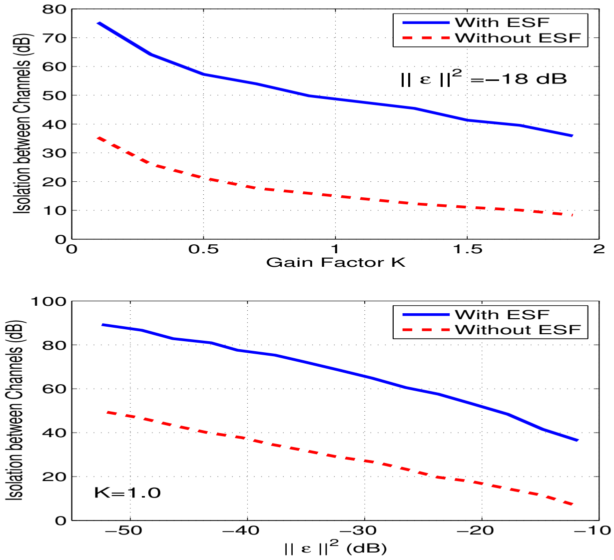


**Fig. 3**. *Isolation between channels with and without ESF in the closed-loop subsystems.*

## 4. PERFORMANCE MEASURES

In this section, a performance evaluation of the echo suppression filters for the closed-loop subsystems is presented.

For the evaluation, we used four different impulse responses corresponding to four different real electro-acoustic paths measured in a medium-size car with 600 coefficients each, using a sampling rate of 8 kHz.

The misadjustment between the impulse response of the electro-acoustic path and the impulse response of the corresponding adaptive filter is controlled by adding a random noise to each one of the coefficients of the original impulse response. This estimation error can be measured by using the normalized $l_2$ norm of the weight misadjustment vector defined as

$$\|\epsilon\|^2 = \frac{\sum_{k=0}^{L} \left| h'_k - \hat{h}'_k \right|^2}{\sum_{k=0}^{L} |h'_k|^2}, \quad (18)$$

where $h'_k$ is the kth coefficient of the impulse response of the real electro-acoustic path and $\hat{h}'_k$ is the kth coefficient of its corresponding adaptive filter.

Several noise free speech recordings were used as passenger's speech adding real car noise, recorded while driving on a highway, as background noise resulting in a SNR around 20 dB. The length of each signal frame was 16 ms and to reduce the overall delay of the system and the time overlap was 75%.

In order to measure the benefit of using the ESF for the closed-loop subsystems, the isolation between channels

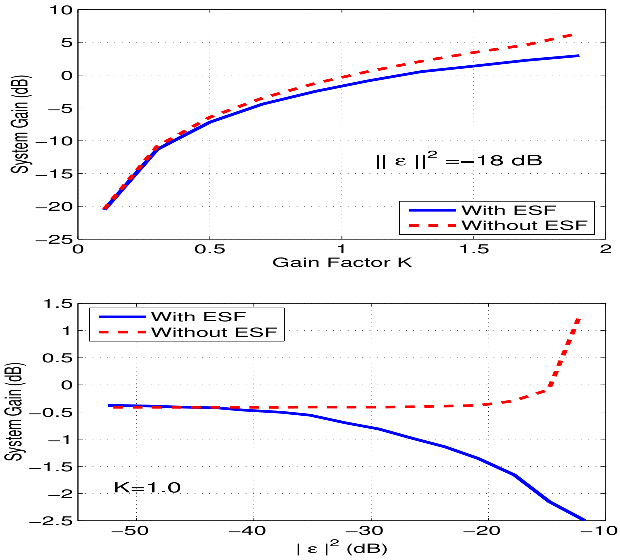


**Fig. 4**. *System gain with and without ESF in the closed-loop subsystems.*

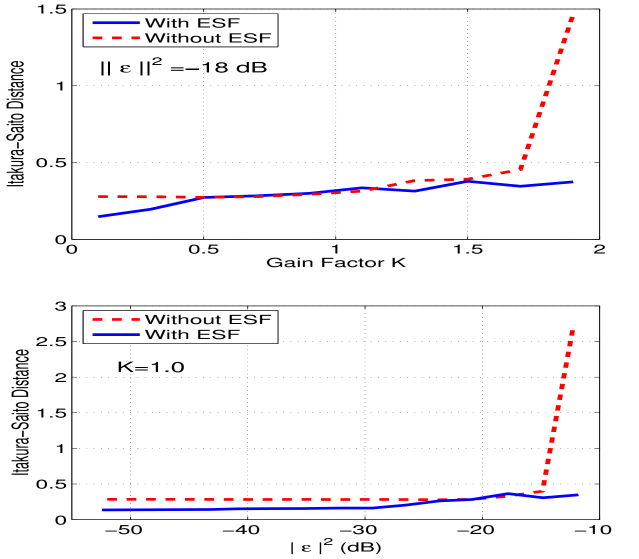


**Fig. 5**. *Itakura-Saito Distance between the input signal and the reinforced speech with and without ESF in the closed-loop subsystems.*

is used. That is defined as the ratio between the power of the front-rear channel output and the power of the rear-front channel output when only the front passenger is talking.

$$I_{RF} = \frac{E\left[|x_R(n)|^2\right]}{E\left[|x_F(n)|^2\right]}. \quad (19)$$

In Fig. 3, the evolution of the isolation between channels with the gain factor for $\|\epsilon\|^2 = -18dB$ is presented. It can be seen that the increase is around 40 dB for almost every value of $K$. The evolution of the isolation between channels with $\|\epsilon\|^2$ is plotted below for $K = 1.0$. The isolation increase ranges from 30 dB for high values of misadjustment (around -12 dB) to 40 dB for lower values of $\|\epsilon\|^2$.

To show that there is no degradation in terms of system gain decrease or distortion increase, the evolution of the system gain with $K$ for $\|\epsilon\|^2 = -18dB$, and the evolution of the system gain with $\|\epsilon\|^2$ for $K = 1.0$ is presented in Fig. 4. In Fig. 5, the evolution with $K$ for $\|\epsilon\|^2 = -18dB$ and with $|\epsilon\|^2$ for $K = 1.0$ of the Itakura-Saito distance between the input signal and the corresponding output signal is depicted.

In Fig. 4 it can be seen that the System Gain increases dramatically for values of $\|\epsilon\|^2$ above -15 dB. The same effect can be observed in both parts of Fig. 5 regarding the distortion for high values of $K$ or $\|\epsilon\|^2$ . This is due to the appearance of howling as the system is very close to instability and strong tonal components are present in the output signal.

## 5. CONCLUSIONS

In this work a two-channel speech reinforcement system has been presented. This system is required in order to make communications inside a car more comfortable. In a two-channel system, two subsystems can be distinguished for each channel, an open-loop and a closed-loop subsystem. The use of specific treatment for residual echo attenuation in the closed-loop subsystems has been presented, and the optimal expression for the transfer function of the Echo Suppression Filter that ensures unconditional stability has been derived. Optimal Echo Suppression Filters do not always exist and the existence of the optimal filters depends on the misadjustment function between the electro-acoustic path impulse response and the adaptive filter of the acoustic echo canceller which is not under the control of the designer. An alternative solution based on an estimation of the residual echo power spectral density is proposed and evaluated. The performance evaluations show that there is an increase of around 40 dB in the isolation between channels when using the proposed Echo Suppression Filters, without decreasing the gain of the system or increasing the speech distortion.

## 6. REFERENCES


[1] Alfonso Ortega, Eduardo Lleida, and Enrique Masgrau, “Speech reinforcement system for car cabin communications,” *IEEE Trans. Speech Audio Processing*, vol. 13, no. 5, pp. 917–929, September 2005.

[2] Eberhard Haensler and Gerhard Uwe Schmidt, “Hands-free telephones - joint control of echo cancellation and postfiltering,” *Signal Processing*, , no. 80, pp. 2295–2305, January 2000.

[3] Gerald Enzner, Rainer Martin, and Peter Vary, “Unbiased residual echo power estimation for hands-free telephony,” in *Proceedings of ICASSP* , May 2002, vol. 2, pp. 1893–1896.

[4] Alfonso Ortega, Eduardo Lleida, and Enrique Masgrau, “Residual echo power estimation for speech reinforcement systems in vehicles,” in *Proceedings of Eurospeech* , September 2003.